

## TITLE OF THE INVENTION

Method To Overcome Instability Of Ultra-Shallow Semiconductor Junctions

---

## CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of the following U.S. Provisional Application:

5 No. 60/397,554, filed July 22, 2002

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR  
DEVELOPMENT

Not Applicable.

## REFERENCE TO A "SEQUENTIAL LISTING," A TABLE, OR A COMPUTER

10 PROGRAM LISTING APPENDIX SUBMITTED ON A COMPACT DISC

Not Applicable.

## BACKGROUND OF THE INVENTION

## FIELD OF THE INVENTION

The present invention relates to microelectronics and, in particular, to a  
15 method of fabricating an ultra-shallow junction in Field Effect Transistor (FET)  
devices, such as Bipolar and Complementary Metal Oxide Semiconductor  
(CMOS) devices. As used in this patent, an "FET" comprises a micro-  
electronic device having a source region and a drain region formed in mutually  
spaced adjacency in the surface of a semiconductor substrate, having a pair of  
20 the shallow implant regions being disposed between and formed as spaced-  
apart extensions of source and drain regions to form a channel region between  
the spaced-apart shallow implant regions, and a gate electrode overlying the  
channel region.

## DESCRIPTION OF RELATED ART

25 Advances in the miniaturization of CMOS devices have been a key  
driving force behind the explosive growth of various network centric computing

products. Smaller CMOS devices typically equate to faster switching times that lead to faster and better performance.

The process of miniaturizing CMOS devices involves scaling down various horizontal and vertical dimensions of the CMOS device structure. In particular, the thickness of the ion implanted source/drain junction of a p- type or n-type transistor can be scaled down with a corresponding scaled increase in the substrate channel doping. Continued scaling of silicon device dimensions down to sub-100 nm dimensions requires highly doped ultra-shallow junctions.

The formation of source/drain extension junctions in CMOS devices is commonly carried out by the implantation of ions in appropriately masked source/drain regions of a silicon substrate with boron (p-type) or arsenic and phosphorous (n-type) dopants. However, ion implantation also creates extensive crystal damage and excess silicon interstitials. Silicon interstitials are displaced silicon atoms created by ion bombardment of the crystalline silicon substrate. In order to activate implanted dopants and remove implantation-induced damages, high temperature annealing at temperature, typically in the range between 800°C and 1200°C, is needed. During thermal annealing, however, the presence of these excess silicon interstitials greatly enhances (10 to 1000 times the normal) diffusion of dopants through the silicon substrate, and results in a much deeper source/drain junction and a poorer junction profile.

In order to increase dopant concentration and minimize boron enhanced diffusion, advanced annealing techniques like spike annealing or impulse annealing currently are used. The strategy behind these annealing methods is to expose the samples at the peak temperature for little or no dwell time. A patent by Chuang entitled "Method to Improve Resistance Uniformity and Repeatability for Low Energy Ion Implantation", U.S. patent 6,362,081, discloses using a spike annealing process for recovering the crystalline structure of the damage silicon layer which results from the low energy ion implantation process. The full disclosure of U.S. Patent No. 6,362,081 is

incorporated into this patent. The rapid thermal annealing (RTA) is executed by rapidly heating the silicon layer to a specific temperature and then instantaneously lowering the specific temperature to a room temperature. The specific temperature is 1100°C.

5           However, stability of those junctions formed by spike annealing is very poor. Dopant stabilization is important to device performance. If spike annealing cannot repair the implantation damage completely, or if new interstitial sources are generated during spike annealing, the interstitials will cause the dopants to redistribute, and deepen the junction depth during  
10           subsequent low temperature processes.

Instability is also an issue for junctions fabricated by alternative methods. A patent issued to D.J. Eaglesham and H.-J. Gossmann, entitled "Forming A Semiconductor Layer Using Molecular Beam Epitaxy," U.S. patent 5,169,798, discloses producing junctions by molecular beam epitaxy (MBE)  
15           growth. The full disclosure of U.S. Patent No. 5,169,798 is incorporated into this patent. It has been shown that junctions formed by MBE are not stable during subsequent thermal processes. Such instability reduces the potential of MBE growth as an integratable method for junction formation due to necessary high-temperature post-growth steps such as ohmic contact formation. For  
20           example, TiSi<sub>2</sub>, a predominant silicide used for interconnects to CMOS devices, requires a temperature higher than 700°C to form a low resistivity, stable phase.

Other related art is disclosed in the following references, all of which are incorporated into this patent by this reference:

- 25           1. Lin Shao et al. " Stability of Ultra-Shallow Junction Formed by Low Energy Boron Implant and Spike Annealing". J. Appl. Phys. 92, 5788(2002).
2. P.E. Thompson and J. Bennet, "Formation and Thermal Stability of Ultra-Shallow p<sup>+</sup> Junctions in Si and Si<sub>1-x</sub>Ge<sub>x</sub> Formed By Molecular Beam Epitaxy", J. Appl. Phys. 92,  
30           6845 (2002).

3. Aliette Mouroux, et al., "Phase Formation and Resistivity In The Ternary System Ti-Nb-Si", J. Appl. Phys. 86, 2323.

What is needed is a solution to overcome the instability of semiconductor junctions.

5

#### BRIEF SUMMARY OF THE INVENTION

The present invention provides a method of forming a microelectronic structure on a semiconductor wafer having a silicon surface layer on a substrate comprising the following steps: implanting dopant ions into the surface layer; cleaning and oxidizing the surface layer, subjecting the structure to a first rapid thermal annealing process, including a first cooling process; subjecting the structure to a second thermal annealing process, including a second cooling process.

The first annealing process, accomplished via rapid annealing, comprises subjecting the structure to thermal annealing with temperature and time of the annealing, as well as the heating and cooling rates, selected such that a majority of dopants are electrically activated, and a majority of substrate damage is repaired. The second annealing process is performed with temperature, duration, and heating and cooling rates selected such that minimal dopant diffusion occurs.

The first annealing process uses a temperature range of 800°C to 1200°C for a duration from about a fraction of a second to less than about 1000 seconds, with a ramp-up rate of about 50°C/second to about 1000°C/second. The second annealing process uses a temperature range of 400°C to 650°C for a time period of from about 1 second to about 10 hours, and more preferably, from about 60 seconds to about 1 hour.

An advantage of the present invention is that it can be used to make structures that are useful in fabrication of microelectronic devices, such as FET or CMOS devices, with significantly enhanced stability during followed thermal processes.

## BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Fig. 1 is a graph that plots boron concentration (atoms/cc) versus depth (Angstrom), depicting SIMS profiles of boron in as-implanted and annealed samples.

Fig. 2 is a graph that plots boron diffusivity versus annealing time (seconds), depicting diffusivities extracted from Fig. 1, and from SIMS profiles of samples annealed under other temperatures.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method of fabricating microelectronic structures and devices having a stable ultra-shallow junction. The junction may have a depth of about 10 to 60 nanometers.

As an initial step, dopant atoms are implanted onto the surface of semiconductor substrate. This implantation can be accomplished in various ways. In one embodiment, a single low energy ion implantation step is employed. In another embodiment, multiple implantation steps may be employed, with each implantation step being accomplished at a different energy level. For example, one may utilize a high energy ion implantation step followed by a low energy ion implantation step. High energy implantation is carried out at energy levels from about 200 kiloelectron-volts (keV) to 2000 keV. In this example, suitable first ions that may be implanted by high energy implantation include silicon, germanium, arsenic, indium, gallium, and other ions. The dosage of the high energy ion is typically about  $1 \times 10^{13}$  ions/cm<sup>2</sup> to about  $1 \times 10^{16}$  ions/cm<sup>2</sup>.

Low energy energy dopant ions are typically implanted at energy levels from about 50 eV to about 5000 eV. Dosages of low implantation ions can range from about  $1 \times 10^{13}$  ions/cm<sup>2</sup> to about  $1 \times 10^{16}$  ions/cm<sup>2</sup>. The particular identity of the dopant ion or ions for low energy ion implantation depends on the type of junction being formed. For example, if a p-type junction is being formed, the dopant ion is usually boron. When an n-type junction is being

formed, the dopant ion is usually arsenic, phosphorus, or antimony. The source of the dopant ion is usually in the form of a molecular complex ion derived from  $\text{BF}_2$ ,  $\text{B}_{10}\text{H}_{14}$  or  $\text{As}_2$  gas sources.

After conducting the implantation, the semiconductor material is subjected to a first annealing process under conditions that cause substantial annealing of substrate damages, for a period of time sufficient to activate dopants. A rapid thermal annealing process with high ramp-up and ramp-down rates is employed. This first annealing process can be carried out in any non-oxidizing environment such as a vacuum, nitrogen gas, or an inert gas. In a preferred embodiment, the semiconductor material is heated in an inert gas atmosphere from room temperature to a temperature of about  $800^\circ\text{C}$  to about  $1200^\circ\text{C}$ , with a ramp-up rate of about  $50^\circ\text{C}/\text{second}$  to about  $1000^\circ\text{C}/\text{second}$ . In another preferred embodiment, the semiconductor material is heated from room temperature to a temperature of about  $900^\circ\text{C}$  to about  $1100^\circ\text{C}$  with a ramp-up rate of about  $50^\circ\text{C}/\text{second}$  to about  $500^\circ\text{C}/\text{second}$ .

After the desired temperature is reached, the temperature is held for a time period of from about 1 millisecond to about 1000 seconds. Once the desired holding time has elapsed, the semiconductor material is then cooled down to room temperature at a ramp-down rate from about  $50^\circ\text{C}/\text{s}$  to about  $1000^\circ\text{C}/\text{s}$ . In a preferred embodiment, the cooling rate is from about  $50^\circ\text{C}/\text{s}$  to about  $500^\circ\text{C}/\text{s}$ .

The above-described RTA is carried out using an apparatus containing, for example, a halogen lamp, an arc lamp, a  $\text{CO}_2$  laser or a microwave cavity or gyratron or an intense hot gas stream. A stationary furnace containing prescribed heating zones could also be used, provided that the furnace contains a means for transferring the semiconductor materials along the different heating zones.

After the first annealing step, the semiconductor material is subjected to a second annealing step. This step can be carried out in any non-oxidizing environment such as a vacuum, nitrogen gas, or inert gas. In a preferred embodiment, the semiconductor material is heated in an inert gas atmosphere at a temperature of about  $400^\circ\text{C}$  to about  $650^\circ\text{C}$  for a period of time about 1

second to 10 hours. In another preferred embodiment, the annealing temperature is from 400°C to 650°C for a period of time from 1 minute to 5 hours. In the most preferred embodiment, the annealing temperature is from 500°C to 650°C for a period of time from 1 minute to 1 hour. Equipment used  
5 for the second annealing step can be those used in RTA, or, preferably, a stationary furnace. The semiconductor material can be loaded in the heating zone with temperature increased from the room temperature to the desired temperature, or pushed into the hot zone when a processor has obtained desired temperatures. After the annealing, the semiconductor material can be  
10 "naturally" cooled down or pulled out from the heating zone. The ramp-up and ramp-down rate for the secondary step annealing can be 0.01°C/s to 1000°C/s.

The following examples are given to illustrate the scope of the present invention. Because these examples are given for illustration purposes only, the invention should not be limited to these examples.

#### 15 EXAMPLE 1

In this example, an *n*-type Czochralski-grown silicon (100) wafer was implanted at room temperature with boron ions at an energy of 0.2 keV to a dosage of  $1 \times 10^{15}$  ions/cm<sup>2</sup>. The wafer was then annealed at 1100°C for 1  
20 second. Annealing was performed with a commercial AG 210 T rapid thermal processor under flowing nitrogen gas. This system uses tungsten filament lamps in a quartz chamber to perform RTA on 4-inch wafers. A typical ramp-down rate for the RTA system is about 100°C/s. After RTA, samples were then annealed at 700°C to study the thermal stability of junctions. Diffusion profiles  
25 after annealing were measured using secondary ion mass spectrometry (SIMS). After 0.2 keV boron implantation, the samples were spike annealed at 1100°C for 1 sec., followed by furnace annealing at 700°C for 30 seconds.

Referring now to Fig. 1(a), it shows SIMS profiles of boron in the silicon samples, after 1100°C spike annealing and additional furnace annealing at  
30 700°C. Fig. 1(a) shows that 1100°C for one second spike annealing is not stable at 700°C. Anomalous diffusion was observed for the samples annealed

at 700°C for 30 seconds with a diffusion length of around 5 nanometers measured at  $1 \times 10^{18}/\text{cm}^3$ .

Referring now to Fig. 1(b), the inventors added a low temperature annealing between the 1100°C RTA and 700°C annealing. The spike-annealed sample was furnace annealed at 550°C for 30 minutes followed by a 700°C RTA for 30 seconds. Compared with control samples that had not undergone 550°C annealing, the samples that underwent 550°C annealing showed no observable diffusion from SIMS. This indicated that an annealing at 550°C for 30 minutes stabilized the junction.

Figure 2 shows B diffusivities extracted from Fig. 1 and from SIMS profiles of samples annealed under other temperatures. The result clearly demonstrates the effectiveness of the second annealing step to increase the stability of a junction formed by spike annealing. With 550°C annealing, B diffusivities following 700°C was significantly reduced.

This invention can be used to stabilize junctions formed by other methods. For example, the shallow junctions formed by MBE growth or laser annealing face the same issue of instability. The procedure of the present invention can be widely applied to junctions formed by various non-equilibrium processes. Thus, the invention is also a method of forming a microelectronic structure on a semiconductor material by molecular beam epitaxy growth, comprising the steps of:

- a. exposing, in a vacuum chamber, a single crystal semiconductor body to a flux of one or more atomic or molecular species, with the body maintained at a temperature greater than about 100°C and less than about 800°C;
- b. depositing a single crystal epitaxial layer with doped atoms that are electrically active; and
- c. subjecting the semiconductor material to a post-growth annealing process.



More specifically, the invention is a method as described, wherein the post growth annealing process occurs in situ in one selected from the group consisting of: a vacuum, nitrogen gas, and inert gas. Even more specifically, the invention is a method as described, wherein the annealing process  
5 comprises heating the semiconductor material with such temperature, amount of time, and heating and cooling rates so that minimal dopant diffusion occurs.

While this invention has been described fully and completely, it should be understood that, within the scope of the appended claims, the invention can be practiced other than as specifically described. Although the invention has  
10 been disclosed with reference to its preferred embodiments, from reading this description those of skill in the art can appreciate the fact that changes and modifications may be made which do not depart from the scope and spirit of the invention as described above, and as claimed in the following claims.